DOCUMENT RESUME

ED 381 552 TM 022 837

AUTHOR Picus, Lawrence O.
TITLE A Conceptual Framework for Analyzing the Costs of Alternative Assessment.
INSTITUTION National Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, CA.; University of Southern California, Los Angeles.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
REPORT NO CSE-TR-384
PUB DATE Aug 94
CONTRACT R117G10027
NOTE 38p.
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Cost Effectiveness; *Educational Assessment; Educational Change; Educational Planning; Elementary Secondary Education; *Expenditures; Resource Allocation; *State Programs; Test Construction; *Testing Programs
IDENTIFIERS *Alternative Assessment

ABSTRACT

A conceptual framework is developed for analyzing the costs of alternative assessment instruments. The framework will be used in future analyses by the Center for the Study of Evaluation to determine the costs of alternative assessment programs in a number of states. It is important that a distinction be made between costs and expenditures. Costs are seen as a measure of benefits foregone to realize some outcome or benefit, and expenditures are a measure of resource flow regardless of their consequence. With this distinction in mind, the paper goes into the specifics of developing a cost analysis for alternative assessment. The first step is to identify the expenditures necessary to operate the assessment program. Since the expenditure estimate alone may not reflect total costs, other complex cost considerations must be resolved, such as the value of personnel time. Determining the value of opportunity costs will improve the quality of educational cost analyses dramatically. Two figures illustrate the discussion. (Contains 10 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

CRESST

National Center for Research on Evaluation, Standards, and Student Testing

ED 381 552

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

A Conceptual Framework for Analyzing the Costs of Alternative Assessment

CSE Technical Report 384

Lawrence O. Picus

CRESST/University of Southern California

TM 022 837

- UCLA Center for the Study of Evaluation

in collaboration with:

- University of Colorado
- NORC, University of Chicago
- LRDC, University of Pittsburgh
- The RAND Corporation

BEST COPY AVAILABLE

A Conceptual Framework for Analyzing the Costs of Alternative Assessment

CSE Technical Report 384

Lawrence O. Picus
CRESST/University of Southern California

August 1994

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
Graduate School of Education and Information Studies
University of California, Los Angeles
Los Angeles, CA 90024-1522
(310) 206-1532

Copyright © 1994 The Regents of the University of California

This research was supported under the Educational Research and Development Center Program cooperative agreement R117G10027 and CFDA catalog number 84.117G as administered by the Office of Educational Research and Improvement, U.S. Department of Education. The findings and opinions expressed in this report do not reflect the position or policies of the Office of Educational Research and Improvement or the U.S. Department of Education.

1. INTRODUCTION

Student assessment in the United States is undergoing a dramatic change as old models of testing are replaced by alternative forms of assessment. These alternative assessment instruments come with a variety of names including portfolios, performance assessments, computer simulations, demonstrations and others. While our understanding of how each of these assessment instruments can best be used is growing, information on their costs is virtually nonexistent. Research at the Center for Research on Evaluation, Standards, and Student Testing (CRESST) has found that policy makers have little information about the costs of alternative assessments, and that they are concerned about the cost trade-offs involved in using alternative assessments compared to the many other activities they feel continue to be necessary.

This problem is exacerbated in times of fiscal retrenchment. The current economic downturn, which, for the fourth year in a row, has left state and local education budgets feeling the pinch, has led to a new fiscal austerity. One result of this austerity is that new program initiatives are receiving increased fiscal scrutiny. If alternative assessment instruments are to be introduced on a broad basis, it will be important to have detailed knowledge of their costs. An understanding of the costs of alternative assessment programs should include, at a minimum, development, implementation and ongoing costs of various programs, as well as an analysis of who actually bears those costs.

The purpose of this paper is to develop a conceptual framework for analyzing the costs of alternative assessment instruments. Before one can reasonably estimate the costs associated with a new policy initiative, or in this case different forms of assessment, it is essential to have broad-based agreement on the components of those costs and how they can best be estimated. Unfortunately, reaching this agreement is not always an easy task. A number of important issues must be resolved before accurate estimates of costs can be developed. Central among those issues is the development of a clear definition of what constitutes a cost. For example, it is relatively straightforward to put a dollar price on the cost of developing and producing new test items and materials, but how should the costs of additional time spent by teachers in preparing and evaluating student portfolios be estimated if they are not provided additional compensation for this task? Similarly, if a performance assessment is

used to replace a traditional multiple-choice examination, what is the "cost" of the information forgone by not administering the previous test? If both assessment instruments are used to assess how students are performing, what are the costs in terms of lost study time for the second assessment? These and a host of similar issues are addressed in this paper.

No attempt is made here to estimate the actual costs of various alternative assessment instruments or programs. Rather the purpose is limited to establishing a framework that can be used to develop consistent estimates of the costs of alternative assessment regardless of the assessment technique employed or the location where it is used. The focus is to establish a methodology for future analyses of the costs of alternative assessments. This framework will be used in future CRESST research to determine the costs of alternative assessment programs in a number of states across the nation.

There are a number of complex conceptual issues that must be addressed in thinking about the estimation of the costs of alternative assessments. Perhaps the most important, and least understood, is what Monk (1993) refers to as the distinction between expenditures and costs. Because of the importance of this issue in developing a complete understanding of the full costs of any program, including alternative assessments, the second section of this paper is devoted to a discussion of how analysts distinguish between costs and expenditures.

Following this analysis, the paper delves into the specifics of developing a cost analysis for alternative assessment programs in education. The third section enumerates many of the expenditures, or actual costs that must be considered in an analysis of program costs, identifying such items as development, production, training and reporting costs among others. The fourth section offers a discussion of "costs" or, as an economist would say, the opportunity costs of assessment programs. The problem here is that these costs are much harder to identify and measure than the expenditures identified in the second and third sections. For the most part, these costs center on the use of personnel, and the time students spend on assessment tasks. Finally, the fifth section of this paper will summarize the conceptual framework and offer conclusions regarding the development of cost estimates for alternative assessment programs.

2. UNDERSTANDING THE DISTINCTION BETWEEN COSTS AND EXPENDITURES

In his work for the New Standards Project, Monk (1993) suggests that a distinction must be made between costs and expenditures and devotes a great deal of attention to the complex issue of distinguishing between the two. He defines costs as a measure of benefits that are forgone to realize some outcome or benefit, and expenditures as a measure of resource flows regardless of their consequence. While this may seem an unnecessary distinction at first, analysis of expenditures to be made on a project, without a thorough consideration of its true costs may seriously under- or overstate the project's cost. Unfortunately, in educational settings, information about expenditures is more easily available than cost information, which has been more difficult to develop. This section begins with a discussion of expenditures and follows with establishment of a more complete economic definition of costs.

Defining Expenditures

A common approach to comparing the costs of alternative programs in educational institutions is to determine the monetary value of the resources necessary to implement each program, and compare the total expenditures across programs. Monk (1993) points out that this process implicitly assumes the two programs are intended to accomplish the same goals, and that both have identical inefficiencies in their operation. If these conditions do not hold, and there is little reason to expect that they do, comparisons of expenditures are invalid and can be misleading (Monk, 1993).

If, as is often the case in education, there are multiple goals established for an alternative assessment program, then estimation of the costs of that program must include all of the resources necessary to accomplish all of those goals. The difficulty is that a project's goals can be difficult to quantify or may even be contradictory. For example, among the many goals that have been attributed to performance assessment are: to change what is taught and learned in schools focusing more on problem solving and critical thinking (O'Neil, 1992); to raise expectations of students (Herman, 1992); and to motivate student interest and effort in learning (Wiggins, 1992). Determining the resources necessary to achieve each of these goals is, at best, a difficult task. Because of this difficulty,

many analysts stop short of estimating the true costs of a program, and instead focus on the expenditures required for its implementation.

In K-12 educational institutions, even determining the actual expenditures for a specific program can be difficult. Most state accounting systems require school districts to report spending by object (salaries, benefits, supplies, etc.), and sometimes by function (instruction, administration, instructional support, maintenance and operations, transportation, etc.). Often these expenditure data are reported at the district level, and there is little or no information about how funds are used at the school or classroom level.[1] Moreover, detailed information about specific programs within a district is often hard to discern from school district financial reports. In an object-oriented system, estimating the expenditures for student assessment might require determining the salaries and benefits of staff members who work in that program, estimating what portion of their time is devoted to the assessment program, and then determining which of the district's expenditures for supplies and materials (including the tests) should be attributed to the program. These expenditures may be coded in different places in the district's accounting reports, making their estimation more difficult.

Even in districts that are able to provide detailed information about the expenditures made for their assessment program, this information only provides a partial delineation of the full economic costs of the assessment program. The other factors that must be considered when estimating the full costs of a program are described below.

Defining Costs

The textbook definition of the cost of a program is the benefits that are not realized through the best forgone alternative. Thus, if a resource is devoted to some use, the benefits associated with the best possible alternative use of that resource represent the "opportunity cost" of the program. Unfortunately, it is not always possible to determine what the best alternative use of those resources might be. Moreover, if that alternative can be identified, determining its benefits may be a considerable problem. For example, if a district is considering

---

[1] For details on the limitations of school district accounting systems for detailed expenditure analyses, see Picus 1993a, 1993b.

the implementation of a new performance assessment program, the opportunity costs of that program would be equal to the benefits from any conceivable alternative reform that was not implemented.

In analyzing the costs of alternative assessments, the range of alternative programs could be thought of as all the possible alternative programs the district could establish to improve student performance. In this case, the benefits derived from the alternative assessment would be compared to the benefits derived from the best option facing the district other than the assessment program. The more beneficial the alternative given up, the more it will cost to devote resources to performance assessment (Monk, 1993). However, before the benefits of a program can be measured, agreement must be reached as to the goals of the forgone activity. As discussed above, reaching an agreement about the programs goals may be difficult to do.

In some cases, it may be appropriate to restrict the alternatives considered. For example, when analyzing the costs of an assessment program, it may well be that the decision to be made is whether or not to replace the existing conventional assessment system with a new form of assessment. In that case, the relative set of alternatives is limited to the assessment program currently in place, and the costs of the new assessment program will be measured on the basis of the forgone benefits of the old assessment program (Monk, 1993).

It is also unlikely that the new assessment program will require exactly the same level of resources consumed by the old system, and it is also possible that a district would be reluctant to eliminate all of its previous assessment program and shift entirely to a new system overnight. In both cases, the total resources devoted to assessment would need to be increased. In the first case, if the new program replaced the old program in its entirety, but required more resources than the old program, the forgone benefits would include both the forgone benefits of the old assessment program plus all the benefits forgone from other activities due to the transfer of resources to the alternative assessment program.

Similarly, if the two programs were operated together, the costs of the alternative assessment would include the benefits forgone by shifting resources into assessment from other areas. If the old assessment program were only continued partially, then the costs of the alternative assessment program would include the portion of the benefits forgone from the old program, along with any

other benefits forgone through the resources that were shifted to the assessment program.

To make a cost analysis useful to decision makers, cost analysts need to develop a common metric to measure the benefits of alternatives. Unfortunately, there is no simple way to compare the benefits of programs that have disparate goals. Since agreement on the spectrum of benefits may also be difficult to achieve, and since estimation of the benefits of a forgone alternative may require a great deal of time for what could be considered an activity with little value (after all, why calculate the benefits of something you do not plan to do?), many analysts simplify the issue by estimating the expenditures necessary to operate the alternative program. As Monk (1993) points out, this is equivalent to estimating the monetary value of the resources directed toward the new program. In effect, the analyst uses the dollar value of the actual or anticipated expenditures as a measure of the projects costs. Often called the ingredients method (Levin, 1983), this approach relies exclusively on expenditures to measure costs, and as Monk (1993) argues, leads to confusion about the difference between expenditures and costs.

If one believes that the benefits to be derived from an alternative assessment dramatically exceed the system being replaced, or if one anticipates that there will be improvements in student learning as a result of the new assessment system (an argument frequently made in favor of performance assessments and portfolios), then using the expenditures devoted to the alternative assessment program may in fact overstate the true costs of the program since the benefits derived exceed the benefits from the program or programs it replaces. Unfortunately, there is no way to estimate the size of this exaggeration. To resolve this problem, the analyst must make explicit assumptions about what factors could cause this overstatement, and then estimate costs with and without an adjustment for this issue. In the framework that is established below, the ingredients approach to costs is used, and where necessary, adjustments for the potential overstatement of benefits that this could lead to are identified and possible adjustments considered.

3. IDENTIFICATION OF EXPENDITURES FOR ALTERNATIVE ASSESSMENT PROGRAMS

The purpose of this paper is to establish a conceptual framework for estimating the full economic costs of alternative assessment programs. Although an important distinction between expenditures and costs is described above, most analysts invariably resort to estimating the costs of a new or alternative program by using the monetary value of the resources devoted to the selected alternative. Since there appears to be no way to resolve many of the complex issues raised above, a thorough understanding of the expenditures necessary to operate an alternative assessment program is essential to this framework.

This section provides a comprehensive list of the elements needed to estimate the expenditures for any alternative assessment program. It is limited to the direct expenditures that would be incurred in implementing any new assessment program, but does not include the estimation of opportunity costs, which is the topic of the next section. Relying on Levin's (1983) ingredients approach, all of the individual items that are purchased as part of an assessment program must be identified and summed to provide a complete picture of the total costs of the program. In this section, a three-dimensional matrix is used to help identify all of these costs.

The first dimension of the matrix relates to the **components** of the assessment program, and includes such things as the development, production, administration and scoring of the test instruments. The second dimension has to do with the **level** at which the expenditures are incurred. Expenditures may be necessary at any one of a number of levels including the state, school district, school, and classroom, and even the private test market. The third dimension of this matrix deals with the specific **kinds** of items purchased for each component at each level, be it personnel, test materials, computer resources, or travel and food for training sessions.

Figure 1 is a matrix showing these three dimensions of **component**, **level** and **kind** of expenditure. The cost of the resources needed for each of the ingredients of the program can be placed into different cells of this matrix. To get an accurate estimate of the expenditures for any assessment program, it is necessary to identify all of the components, levels and kinds of expenditures that must be made.

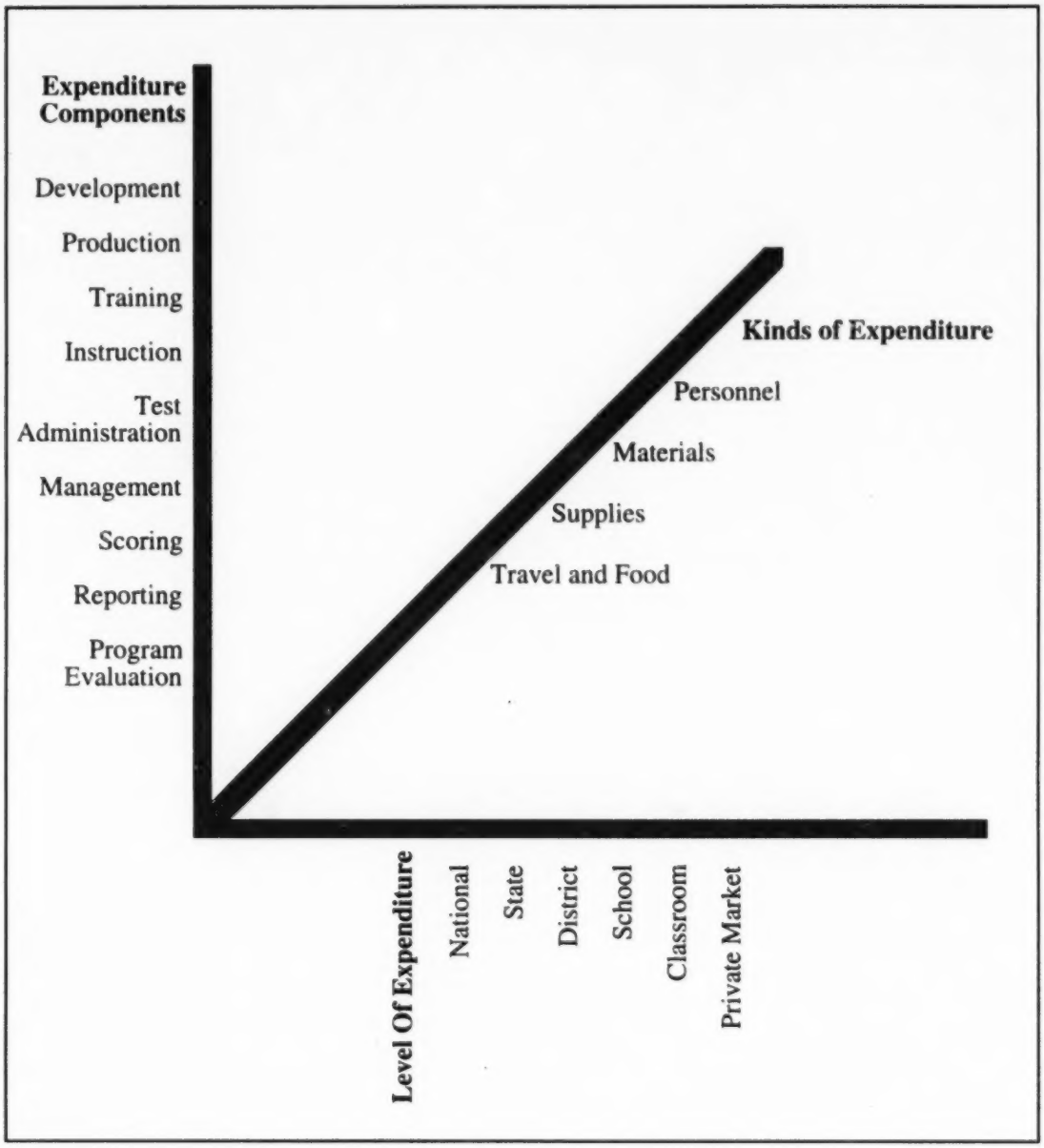

Figure 1. Expenditure dimensions.

As Figure 1 shows, the major expenditure components of an assessment program include the following:

- Development
- Production
- Training
  - For administration of assessments
  - For interpretation of the results
  - In related instructional strategies
- Test administration
- Program management
- Scoring
- Reporting
- Program evaluation

Figure 1 also identifies six different levels where expenditures are likely to be observed, ranging from the national level, to state, school district, school and classroom expenditures. The level of expenditure also includes private test developers who are an important part of the current assessment system, and are expected to be major players in the future. Finally, the figure identifies the kinds of expenditures likely to be made. These reflect the "things" or, in Levin's terms, the ingredients for which dollars are actually spent. The kinds of expenditures include salaries, supplies and materials, and travel and food costs for training sessions.

Below, the expenditure components of an assessment program are described in more detail, and the kinds of expenditures and level where they might be expected to occur are identified. In effect, this discussion will identify the matrix cell from Figure 1 where each expenditure occurs. While the discussion could be based on any of the three dimensions identified in the matrix, the component axis was used here because it seems the most straightforward way to identify the elements that constitute an alternative assessment program.

Development Expenditures

There are two major components to the measurement of expenditures for the development of alternative assessments. The first relates to the establishment of a new assessment program and involve typically one-time costs.

The second component has to do with the development of assessment instruments and has an initial high cost to develop a set of instruments, and then lower costs for development of replacement items over time. Each of these expenditure items is considered below.

Program Establishment

Regardless of the type of assessment program to be implemented, there will be a number of initial development costs that must be measured. For example, if a decision is made to rely on portfolios to assess student performance, then substantial time must be devoted to determining what those portfolios should contain, or, at a minimum, to determining the guidelines individual teachers should use in working with students to create the portfolios. If a performance assessment system is the goal, it will take a great deal of effort to decide what kinds of performance need to be assessed, how performance will be assessed, and how the results will be analyzed and reported.

The level at which these expenditures need to be analyzed depends on where the assessment program is created. To date, efforts to establish a statewide portfolio assessment are underway in Vermont. A number of individual districts across the country are also experimenting with portfolio assessments. Other types of performance assessment are being considered at all levels, from national testing organizations and publishers, down to individual performance assessments created by individual teachers.

State level. If the goal is to understand the expenditures made for the development of an assessment program, it is helpful to begin at the highest level, and work down. If a state testing initiative is the subject of the investigation, then the following kinds of expenditures would need to be considered at the state level:

- *Initial research analyses*. This would include any expenditures incurred to study assessment programs in other states and to estimate the potential value of a new form of assessment. Places where such expenditures might occur include the legislature, the state department of education, and education interest groups working for and against any proposed changes.

- *Feasibility studies*. The state might also incur expenditures conducting feasibility studies to ascertain the level of interest in new assessment methodologies among school professionals, policy makers, and even parents.

- *Design expenditures*. Once a decision has been made to move forward on development of an alternative assessment program, the details of the program have to be fleshed out. This may be done through focus groups, employment of consultants, and by convening teachers and other interested parties to discuss the methods and uses of any new assessment program. Expenditures at this stage include personnel costs at state agencies and education interest groups involved in those discussions. The expenditures would include the value of the time of all individuals participating in the discussions and planning. In addition to the personnel costs, any expenditures for travel (to both attend meetings and observe schools in other locations), for consultants, and for the development of materials to facilitate the decision-making process that are paid by the state or by state-level organizations must be enumerated.

In addition to these state-level expenditures, there will be school district and school-level expenditures in the establishment of the program.

District level. In the instance where the state initiates the assessment reform, the expenditures made by a district would largely be those made through participation in the design and testing of the program. In a district-developed program, the expenditures identified above for the state would be borne instead by the district. The following district expenditure categories need to be considered regardless of whether the alternative assessment program is a state or district initiative.

- *Initial research analyses*. As discussed above, the first step is feasibility studies to ascertain what kinds of assessment are needed in the district, and what is necessary to adopt different assessment options. Focus groups of teachers, administrators and parents may be helpful in learning what kind of assessment is desired, and the employment of outside experts may be necessary. While the bulk of the costs at this stage would appear to be related to the research effort, if the participation of school teachers and administrators is desired, the expenditures required for having them attend meetings must be considered.[2] The primary expenditure is likely to be for teacher substitutes in the case where teachers are involved. Another major cost is for the travel and meals for people to participate in the meetings. This expenditure would be substantially reduced in a district effort since travel distances are likely to be less significant, and overnight lodging unnecessary for teacher from that district.

---

[2] The opportunity costs of the time these individuals spend on this project will be discussed in the next section of the paper. Here we are only concerned with the actual expenditures made by the district to facilitate individual participation in the design and development of the assessment program.

- *Project design*. Once an assessment strategy is selected, the district must put together a design team to develop the specifics of the assessment program. Expenditures at this stage of the venture include costs for consultants, as well as the costs of district personnel who participate in the process. If the work is done during the school year, it may be necessary to pay for substitute teachers when teachers on the design team need to attend other meetings. If much of the work is done during the summer months, then funds must be set aside to compensate teachers and other district staff for their time. Finally, as the design takes shape, and assessment instruments are actually developed, there will be more expenditures for supplies and materials.

Development of Assessment Instruments

The second component of development expenditures relates to the creation of assessment instruments. The type of assessment considered, as well as the number of assessment items necessary, will have a major impact on the costs of development. There is not necessarily a direct relationship between the desirability of the assessment instrument and the cost of its development.[3] However, it does seem logical that the more complex the assessment instrument, the greater the cost of its development. For example, design of a computer simulation is probably a more complex task than developing a short-answer or multiple-choice test item. Once a simulation activity has been "thought through" and the structure designed, considerable time is required to actually program the simulation activity. The time, and hence expenditures, necessary to develop the simulation will increase if complex mechanisms to measure and track individual student performance are included.

Another important expenditure consideration in the development of assessment items relates to the use of manipulatives as part of the assessment. For example, Shavelson and Baxter (1992) describe a mystery box investigation to assess science skills. Clearly having a set of mystery boxes for each student is very expensive. If only one set of boxes is provided, then it is very time consuming to have each student participate in the assessment, particularly if monitoring of the students while they attempt to complete the tasks is essential to a complete assessment. In addition there are risks that the students who

---

[3] There is probably a direct relationship between the total cost of each assessment technique and its desirability, but remember the focus here is only on development expenditures. The expenditures for scoring and reporting the assessment results, as well as estimates of the costs for teacher time, are considered below.

participate in the activity early on will inform their friends as to how to succeed on this task. Paper-and-pencil alternatives are described by Shavelson and Baxter, but they find them to be less valuable than the hands-on experience.

Another issue related to the development of alternative assessment instruments has to do with the number of items necessary to accurately measure a student's knowledge or skill. The more items required, the higher the cost. Shavelson, Gao, and Baxter (1993) found that to establish a generalizable estimate of a school's mean science achievement (regardless of other schools' performance) it would be necessary to sample 50 students on 15 tasks or 100 students on 12 tasks. To rank order schools on the basis of science achievement would require 25 students per school and 10 tasks or 100 students and 5 tasks.

These findings have important implications for both the development and administration expenditures of alternative assessment programs. While there is a considerable trade-off between the number of students tested and the number of tasks to be administered, it seems that in the long run it is less expensive to develop more tasks and test fewer children. Regardless, assuming an absolute score is desired for each school, a minimum of 12 tasks would need to be developed. Moreover, this only provides information about one level of science achievement. The number of tasks would also have to be multiplied by the number of subjects to be assessed, and the grade levels at which the assessments would be conducted. The number of tasks would therefore grow geometrically from the base of 12 identified above.

The development costs for these activities are not well documented to date. To estimate the expenditures necessary to develop individual assessment items would require determining the amount of time spent developing, writing, pilot testing, and evaluating each item, as well as estimating the costs of materials necessary to develop the items. Initially this may be very expensive if a large number of items are to be developed. However, once the initial development stage is complete, the challenge of developing replacement items should be lower than the initial development. As a result, development efforts will need to be maintained, but at a lower level than initially. What this means for a cost analysis is that both initial item development expenditures and long-term development expenditures must be considered in the analysis.

Another important component of the development costs is pilot testing. State-level costs for pilot tests would include the production of assessment instruments, the administration of the assessment, and the costs of analyzing the results and modifying the assessment instruments based on those results. The largest kind of expenditure would be for personnel to administer, score and evaluate the assessment. Many of these expenditures would be for state personnel and/or consultants. In addition, the state might elect to pay for local school personnel either to participate in the pilot test directly, or to help with the administration, scoring and evaluation of the pilot test.

A district participating in a pilot test will have to devote administrative and teacher time to the administration of the examination, which will take away from other activities. The opportunity costs of these activities will be discussed below, but there are some direct costs which must be considered. These include any overtime that the district pays, any substitutes for teachers who participate in training sessions (unless the state pays for this, in which case the expenditures would be identified above), and any expenditures incurred in the evaluation process.

Other cost items to be considered at this stage include development and pilot testing of scoring schemes and the training for scorers. Development of the scoring schemes are largely a one-time task, and the costs of developing training sequences and materials for those sessions are one-time costs. There may be a need to modify scoring schemes somewhat as new assessment instruments are developed, and there will be a recurring need to train new scorers over time due to attrition.

The costs of developing assessment items will be found at the level initiating the assessment program. For example, a state performance assessment initiative will have to account for the expenditure for development of assessment items. Likewise, a district effort will have to factor in the expenditures necessary to develop items for a district initiative. If individual schools or teachers attempt to establish their own alternative assessment programs, then the expenditures necessary for developing items may occur at that level. Even if the district, school, or individual teacher elects to use already developed assessment items, someone will have to actually determine which ones are to be used, and how they will be used. If an individual is compensated directly for this task, as may well be the case in a school district, then the direct

expenditures must be considered here. On the other hand, if individuals make these selections without compensation, then the development would fall into the cost rather than expenditure category. Identification and measurement of these costs is discussed in the next section.

A final level that must be considered here is the private market. Given the intense interest in alternative forms of assessment across the nation, most of the existing test developers are undoubtedly devoting considerable resources to the development of new assessment items. Obviously, it is their hope that these expenditures will be recouped through the sales of the assessment instruments to schools in the future. It may be possible to identify the direct expenditures made by these private firms to develop new items, particularly if such development is done on contract for a public school system, such as the development efforts taking place currently in Kentucky. However, if the development of assessment items is moving forward on the basis of speculation on the part of the private test publishers, then the costs of forgone options must also be evaluated.

Summary

There are two major components to the measurement of expenditures for the development of alternative assessments: (a) expenditures for the initial design and implementation of the program; and (b) expenditures for the development of assessment items. The components of these expenditures are largely for personnel time to plan, design and implement the program, as well as to develop assessment items. These expenditures can occur at any of the levels identified on the horizontal axis of Figure 1, depending on where the initiative for the new assessment program occurs. To the extent that individuals participating in the development of a new assessment program are directly compensated for their time, the expenditures would be identified here. However, to the extent that the development of an alternative assessment program relies on individuals whose time is not directly compensated, the costs of that time must also be considered. Those costs are discussed in the fourth section of this paper which focuses on the measurement of such costs.

Production Expenditures

Once an alternative assessment program has been developed, and a decision is made to use it more generally, there will be expenditures necessary for the production of assessment items. These expenditures are relatively straightforward to measure and are dependent on the number of students to be assessed, the type of assessment to be conducted, and the number of assessment items to be used. The level at which the expenditures occur will depend on the level initiating the assessment program, but is most likely either a state or a school district. Private test publishers may produce many of the assessment items that are actually used, but the most direct measure of the expenditures necessary for the assessment program will be the money spent by the education agency actually purchasing the assessment items from the publisher.

Another important item to consider for forms of assessment that rely on "hands-on" evaluation instruments is whether or not the item itself is reusable. For example, in a science assessment using the mystery boxes described by Shavelson and Baxter (1992), it is likely that the boxes themselves could be used by a number of students, in a number of classrooms and even schools. One factor that would affect the number of mystery boxes needed would be how sensitive assessment is to timing. If all students had to be tested in a brief window of time, more boxes would be required than if the assessment could take place over the course of a semester. The issue to be resolved is the trade-off between total expenditures and the validity of the assessments if they take place over a longer time frame across a school, a school district, or even a state.

Since paper-and-pencil assessments are relatively inexpensive to produce, it is possible to test all students on the same day, or in the same week. However, to do so with a hands-on assessment instrument requires substantially more assessment instruments to be available. There are clearly benefits to having all of the assessments take place at the same time, or within the same brief time frame. However, the expenditures necessary to achieve this may be prohibitive. Figure 2 shows the relative trade-offs between the forgone benefits of a longer time frame and the costs of producing additional assessment instruments. The point where the two cost curves intersect represents the equilibrium between expenditures for the assessment instruments and the forgone benefits of increasing the testing time frame. Identifying this equilibrium would indicate the most cost efficient point for producing assessment items.

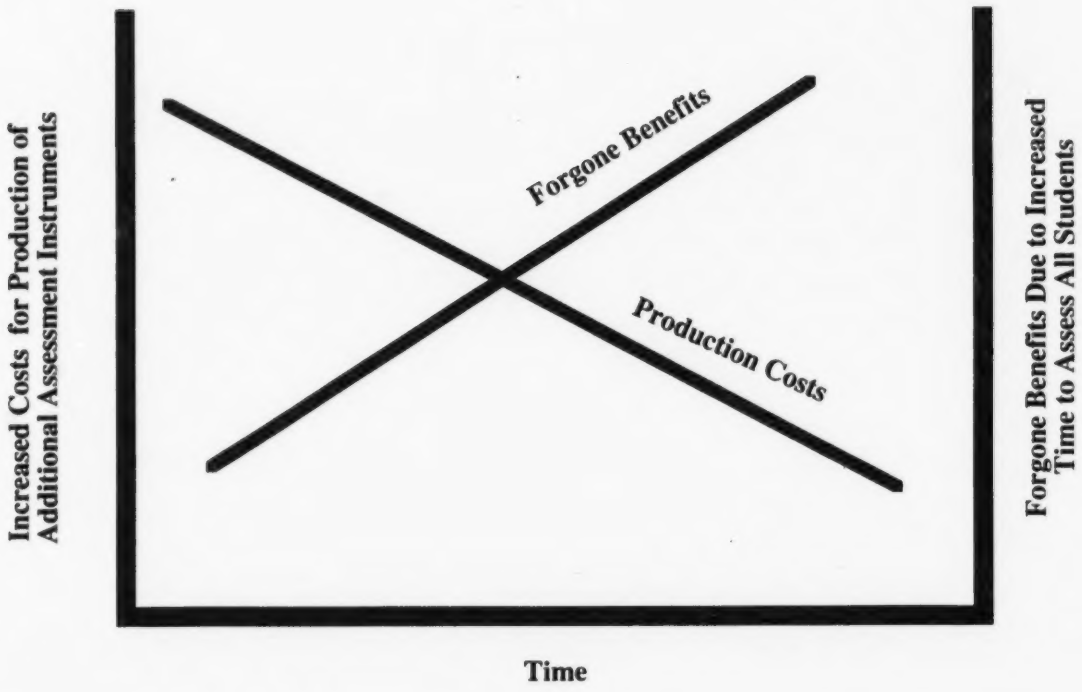
Increased Costs for Production of Additional Assessment Instruments
Forgone Benefits
Production Costs
Forgone Benefits Due to Increased Time to Assess All Students
Time

Figure 2. Trade-off between production costs and increased assessment time.

In summary, production expenditures are relatively straightforward to measure since they relate only to the production of the assessment instruments to be used. However, if the assessment instruments used are expensive to produce, there are a number of trade-offs that must be considered in terms of the number of items produced and the length of time over which the assessments take place.

Training Expenditures

Training expenditures can be divided into three categories: (a) training for the administration of the assessment; (b) training individuals to score, interpret, and use the results; and (c) training personnel in related instructional strategies linked to the assessment's goals. Some of these expenditures can be directly attributed to the implementation of an alternative assessment program, while others could be considered part of the instructional program. Distinguishing between them is relatively straightforward once the actual expenditures have

been identified. Identification of those expenditures is the focus of this section, while the division of them between instruction and assessment is left to individual cost analysts.

Training for Administration of the Assessment

Administration of an assessment can best be thought of as establishing the conditions under which students are to perform. A traditional paper-and-pencil multiple-choice test would require very little training for the individuals whose responsibility it is to make sure that the test is administered properly and at the correct time. If particularly stringent security measures are required, some training for personnel may be required, although it is possible this could be managed through written instructions. Similarly, if teachers or school administrators are required to select a sample of children to test, some training in the selection methodology may be required, although most of these decisions are made by the level centrally responsible for the administration of the test. Finally, some minor training may be required to explain how batch header sheets or similar forms should be completed. In most cases, this training could be accomplished through written instructions. If on-site training were required, there would be expenditures for the time and travel of an individual to go to school sites to explain procedures to site administrators and teachers.

Training to Score the Assessment and Evaluate the Results

Training individuals to score or evaluate the results of an assessment is much more complex than training for the administration of the assessment, and therefore, these costs should be identified separately. As in the discussion above, the costs of training will depend on where the training takes place, how much time is involved, and how many individuals are involved, both trainers and trainees. It is also important to consider training costs both for scoring and evaluating student performance and for understanding the results of the assessment and using those results in developing instructional plans.

If the assessment instrument is more complex, particularly if it requires teachers to evaluate individual student performance in their classrooms through observation, creation of portfolios, or grading of essays, then more training may be required to ensure that students are assessed uniformly and accurately. The complexity of the assessment will determine the training costs to be incurred.

This could range from requiring teachers to read certain materials, to expecting teachers and/or site administrators to attend out-of-town training sessions that lasted for more than one day. Thus the expenditures necessary for the training of individuals to score the assessment will vary with the complexity of the assessment and the amount of effort expected of the individuals actually doing the scoring of the assessment. If it takes 15 minutes to score an individual portfolio, and if multiple raters consider each portfolio, it requires between 30 and 45 minutes of time to assess each student in each subject area, plus the time necessary for training and validation checks.

Training in Related Instructional Strategies

To the extent that a new assessment program is implemented along with or to support change in instructional strategies used in a school or school district, it is important to provide teachers with training in the instructional techniques. Expenditures in this area can rightfully be attributed to the instructional program and not the assessment program. However, to get an accurate portrayal of all the expenditures related to a new assessment program, it is important to identify any of the training costs of a new instructional program that are specifically directed toward the assessment component of the program. As above, these costs will be dependent on the location of the training and the amount of time involved for both trainers and trainees.

Test Administration and Program Management

Identifying the expenditures associated with the administration of an assessment includes the time and materials required to make sure that (a) the assessment materials are delivered to the school sites on time, (b) the instruments are administered during the designated time frame, and (c) the response booklets or other completed instrumentations are delivered to the proper authorities on time, in a secure fashion if necessary. Management of the program includes the tasks necessary to make sure that all of these things happen as well as making sure the evaluation of the program and revisions to the program based on that evaluation are conducted regularly.

The expenditures for administration would include the delivery costs for assessment materials both to and from school sites, as well as the personnel costs associated with making sure that the materials were delivered and picked

up properly. If individuals at school districts or school sites receive compensation for assisting with the administration of the assessment instrument, then those expenditures must also be included in the total. In the case of performance assessments that require specialized equipment or testing materials that are transported from site to site, it is important to include all transportation or delivery expenditures in the total.

Expenditures for the management of the assessment program are likely to occur almost entirely at the highest level. Thus, a statewide assessment program will be administered though the state department of education, and most of the expenditures required for the ongoing management of the program would be found at that level. Some management expenditures might be found in school districts, but most of the kinds of costs one might think of here, such as local site coordinators, are more likely for administration of the regularly scheduled assessment, and not for long-term management. While it is important to distinguish between these two types of expenditures, it is probably more important to make sure that all of the expenditures for these activities are accounted for properly. As with some of the earlier discussions, the costs of personnel time not directly reimbursed through the assessment program are discussed in the section of this paper dealing with costs rather than expenditures.

Finally, as assessment tasks become more complex, the logistics of conducting the assessment and maintaining records so that the results can be linked to individual students become more difficult, increasing costs. Considerable time may be required just for the labeling of answer sheets and assessment booklets, not to mention keeping them sorted properly for shipping to the locations where they will be evaluated.

Scoring

The scoring of multiple-choice tests is a relatively straightforward process which involves the use of machine readers. Because such testing procedures have been in use in educational institutions for many years, there is an abundance of information on the expenditures necessary to score assessments that rely on multiple-choice materials. The expenditures required for alternative assessments are harder to estimate and depend on a number of factors, including

what is assessed, how many individuals need to assess each student's work, and how long each review takes.

The scoring of essay-type examinations takes considerably longer than does the scoring of multiple-choice materials. Moreover, in most cases, multiple readers are employed for each task given. The expenditure for personnel time includes time for training the raters, evaluating interrater validity, and the time it takes to actually review each of the assessment tasks. If there are a substantial number of essay items, it can take a substantial amount of time to complete these tasks. Moreover, if the instruments are scored in a single location, travel and per diem costs must be included in the expenditure total.

A further complexity is introduced if the raters are asked to evaluate student portfolios rather than grade essay examination questions. With essay examinations, each student is asked to respond to the same question or set of questions, and the time spent evaluating each is likely to be relatively consistent. Moreover, once the raters are clear about the objectives of each question they are scoring, consistent scoring is generally possible. However, if each student prepares a portfolio of his or her best works, each portfolio will contain different materials. These differences result both from different assignments by different teachers and from different choices of materials by students in the same class. Making consistent evaluations of such portfolios is more difficult, requiring more time, and possible review by a higher number of raters. Clearly, the more complex the evaluation instrument, the more time will be required to get consistent scoring results across the region participating in the assessment.

As hinted above, the level at which the assessment takes place will have an impact on the total expenditures. The larger the region assessed, the greater the costs of evaluating the student products, in part because the challenge of reaching consensus on standards and score values is more difficult. Moreover, if more than one school district is involved, it is likely that there will be considerable differences in what students were taught or how it was taught, resulting in greater differentiation in student materials. Consistency of scoring across a state, thus, is generally more difficult than similar scoring within a district. In the same way, the costs of assessing student performance within a school may be reduced because of similar consistencies within a school that are not found at the district level, particularly in a large school district. Since it is

likely that individual teacher differences also exist, even within a school consistency issues are likely to arise. Therefore, it is possible that the higher the level of the assessment program, the greater will be the expenditures necessary to score student materials. It is hoped that this idea can be tested empirically in future research.

Assuming that the alternative assessment program relies on materials other than multiple-choice tests, the bulk of the expenditures for scoring student outcomes will be for personnel time to evaluate the materials. There may be expenditures for travel and per diem if raters are brought to central locations either for training or for the actual evaluation of student work. In addition, some computer resources may be necessary to estimate interrater reliability and to keep track of each student's score. There are also costs associated with the logistics of copying and assembling scoring "bundles" and transporting student assessments to the scoring sites.

Reporting

Reporting the results of an assessment to the participating schools, children, parents, and community is an important component of any assessment program. If the assessment program relies on portfolios, or a similar instrument that leads to different materials being submitted by each student, then resources will be necessary to develop comparison across schools and school districts. The expenditures necessary for doing this will depend on the level of comparison desired. As pointed out above, if the goal is simply to be able to rank schools on the basis of performance, rather than develop an absolute score for each school, it may be possible to reduce the number of assessment items, and consequently reduce the costs associated with the scoring and reporting the results.

Another issue has to do with comparing assessment results nationally or regionally. There seem to be two approaches to developing assessments that are comparable across large numbers of states. One is to require a national assessment system that offers the same tests to all students. The other is to develop clusters of tests in different geographical areas, each of which is designed to meet local needs, but which assess the same set of student skills. The results could then be compared across states. While the first of these two options is probably less expensive to develop, the costs of getting all states, not to mention all school districts, to participate is very high. On the other hand,

states may be more willing to participate in examinations that are developed locally, but the costs of developing tests where the results can be compared across states or geographical regions will be considerably higher.

Another comparability issue has to do with comparing the results of any new assessment instrument with the data that were provided by the system it replaces, or along side of which it continues to operate. If the two are to be operated in parallel, then the expenditures on the alternative assessment system will be considerably higher than if the alternative is to replace the existing system. However, the total costs of the system are not as easy to discern. In many schools, school districts, or states, there has been a considerable investment in the existing assessment system. To replace it would require adapting to new kinds of information about students, and giving up information that many have become accustomed to using. The value of the information forgone by eliminating an existing assessment program is a cost of any new system that replaces the old. Developing algorithms to inform users of the new system how the results of the new assessment could be interpreted under the previous system may or may not be possible. If it is possible, expenditures to make that information exchange could be expensive, both in terms of the expenditures for staff and materials to make the translation and in terms of undermining some of the things the new assessment seeks to accomplish in measuring student performance.

As a consequence, many new assessment systems may be run in parallel with the programs they are intended to replace, increasing the costs of student assessment generally. Sorting out the complex interrelationships between the two systems, determining the resources necessary for each, and comparing the costs and benefits of the various alternatives in terms of the operation of either or both programs are important components of any cost analysis. It is likely that an alternative assessment program would be phased in over a period of time, and the system it was designed to replace phased out in a similar fashion. A cost analysis would need to take the expenditures of doing this into account, as well as estimate the relative benefits and costs of the trade-offs that are made as the new system is put into place. To the extent that the two systems operate in parallel, the costs of the assessment program will increase.

One of the hallmarks of many of the new, innovative assessment programs has been the efforts devoted to reporting the results. In Vermont, town meetings

were held across the state to discuss the results of the first-year portfolio assessments with parents, and in California, the CLAS communications package includes a video to help explain the test and how the results can be used and interpreted. These dissemination strategies should make parents and community members more aware of the outcomes of the assessments, but maintenance of this level of effort is expensive. Individuals must be recruited, and paid, to conduct town meetings, and the production and duplication costs for a video can be substantial.

The kinds of expenditures most likely to be required for reporting relate to personnel: to determine technically the best way to condition and convey the results; to develop materials that explain the scores and how to use them; and to distribute results to the proper individuals. There are also material costs for the distribution of those materials, but in general, they would seem to be quite low. The level at which these expenditures would occur depends on the level at which the assessment takes place. In addition, some expenditures would be expected at each level below the level conducting the assessment as local officials work with the data they receive to inform the community, parents, students, and other school staff.

Program Evaluation

Regardless of the assessment alternative in use, an important component of the program should be an ongoing evaluation. Understanding the strengths and weaknesses of the assessment instruments and making sure that the information being provided is what is really sought are important parts of any educational program. Therefore resources should be devoted to the evaluation of the assessment instruments and the information they provide.

The level at which this evaluation takes place, or at least from which it is coordinated, can vary, although for economy, the most likely level is, again, the highest one. This will ensure uniform evaluation standards across districts and schools and provide the widest sample frame for assuring the statistical adequacy of the sample. The largest expenditure is likely to be for personnel who conduct the evaluation. Other expenditures are likely for computer resources, and materials and supplies. In addition, expenditures for travel and per diem for site visits may be a necessary part of any evaluation. Finally, the

expenditures necessary to report the evaluation findings must also be considered.

Summary

This section of the paper has provided a detailed description of the kinds of direct expenditures that can be identified as part of any assessment program. The expenditures have been divided into seven components, although as the discussion reveals, making distinctions between some of the areas for some kinds of expenditures can be difficult and subject to differing interpretations. In addition to these seven components, expenditures can be identified at different levels, and for different kinds of resources. The largest single expenditure item in any assessment program seems likely to be personnel. The more complex the assessment instruments, and the more difficult they are to score, the higher the personnel costs are likely to be. The other two large expenditure categories identified above are the costs of developing and producing assessment items, and travel and per diem costs for training.

To this point, the discussion has focused on direct expenditures that can be identified as being part of the costs of an assessment program. The next section describes the cost factors that must also be considered in analyzing the costs of implementing an alternative assessment program.

4. IDENTIFICATION OF COSTS FOR ALTERNATIVE ASSESSMENT PROGRAMS

The previous section examined the types of expenditures that could be attributed to an alternative assessment program. However, as discussed in the second section of this paper, the expenditures made on behalf of a program may not accurately reflect the true costs of that program. It was suggested there that identification of the full costs of a new program have to take into account the benefits forgone from the best alternative to the selected program. The problems with determining these costs are twofold. First, it is hard to identify and measure the benefits of educational programs, and second, differences in the amount of resources needed for each program alternative can lead to difficulties comparing benefits across programs.

The discussion in section 2 suggests that to get a more accurate portrayal of the true costs of an alternative assessment program, it is necessary to measure both the direct expenditures, as discussed in section 3, and the costs of any other resources that are dedicated to this process, but for which direct expenditures are not made. The most obvious example of these kinds of costs are the opportunity costs for the time individuals spend working on the assessment program instead of on other priorities of the state, district or school.

This section describes the various areas where these opportunity costs might be found and provides some guidance for estimating their costs. Rather than organize the section along the lines of the expenditure components as was done above, this discussion is based on the kinds of costs that might occur, identifying the areas where opportunity costs are likely to be found. Since the largest of these costs is probably personnel time, the discussion begins with an analysis of the costs for personnel.

Personnel Costs

In most K-12 educational institutions, expenditures for personnel represent the largest single spending category. In section 3, personnel costs were identified as the largest expenditure component in most alternative assessment programs. The costs for personnel that are borne by the project itself through direct payments to either staff, consultants, or private concerns are what were identified as expenditures above. If an individual works on parts of the assessment program without direct compensation, then the costs of that person's time must also be accounted for. The difficulty is determining who actually bears that cost.

An example here may be helpful. Suppose that a school district implements a portfolio assessment system for all middle school math classes. Teachers are responsible for putting together a portfolio for each student. The portfolio is supposed to show examples of the student's best work and worst work, and provide some sense of how well the student understands the concepts that are part of each course. Teachers are expected to consult with the students in deciding what should be placed in the portfolio, and some materials will be selected by the student, some by the teacher and student jointly, and some by the teacher alone. Putting together a portfolio of this nature for a large number

of children (suppose an average middle school math teacher sees 20 students in each of 5 classes a day for a total of 100 portfolios) can be very time consuming.

If the teacher receives no additional compensation for doing the work entailed in putting the portfolios together, the time spent on that task will have to come at the expense of something else. The costs of that time will be borne by different levels of the system depending on how each teacher chooses to allocate his or her time given this new, and demanding, task. If the teacher spends an additional two hours a week working on the portfolios, at the expense of his or her leisure time, then the costs of the teacher's time will be borne by the teacher him- or herself.

If the teacher was previously active in a school site-based management committee, or somehow participated in the day-to-day operation of the school, but reduced this participation by the amount of time it took to create the portfolios, the school would bear the cost of the time it takes the teacher to prepare the portfolios. If the teacher uses class time to prepare the portfolios, choosing to provide the students with time in class to do assignments that were previously assigned as homework, and limiting in-class instruction time, then the students will bear the cost of the new assessment program.

Finally, if the teacher does not want to change his or her allocation of time to these various activities, he or she may elect to spend less time putting together the portfolios than is needed to do an adequate job. In this case, it is harder to assign the cost since it represents time not spent, rather than a reallocation of time. If the portfolios provide valuable assessment information about students, which might help them improve their school performance, the students may bear part of the costs of the teacher's decision. If the teacher's future evaluations suffer because of this reluctance to participate, the teacher may bear part of the cost (although given current salary systems in public schools, this cost will be borne through non-monetary means, not reductions in salary growth). Potentially the school could bear part of the cost if it is identified as a low-performing school based on the quality of the portfolios submitted. Finally, the assessment program itself could bear the cost if it is deemed unsuccessful and its size and scope are reduced in the future.

The difficulty here is both measuring the costs and determining who bears those costs. The best way to evaluate the costs of the teacher's time for

developing the portfolios is to use the hourly wage the teacher receives from the school district. While this may not truly reflect the cost to the teacher, as the benefits of additional leisure time or more participation in school governance may be more rewarding to some individuals, it does provide a common metric for placing a value on the time required of all participants in the system.

The second issue is more complex. Assigning the cost to its proper location requires identification of how the teacher changed his or her allocations of time and over what period of time those changes were in effect. Moreover, there is no reason to expect that teachers will all make the same choices. The best way to get at this information is to query individual teachers to see how they modified their activities to meet the new requirements of the portfolio assessment system.

The time teachers devote to an assessment system does not represent the only personnel costs that must be considered. Public school officials, from the state department of education down to the classroom, are likely to have time commitments related to the new program that are not directly compensated. As a result of these new responsibilities, each of these people will have to make decisions as to how they will allocate their time to the new system and to their other responsibilities. Again, it is possible to determine the value of this time by using the hourly wage paid to each person. The more difficult problem is ascertaining how much of this time is paid for directly through the assessment program (which would be treated as an expenditure), and how much comes at the cost of reductions in some other activity. The locus of that cost—the agency or individual who actually bears the cost—must also be determined. The best way to do this appears to be surveys of individuals involved in the process similar to those used by Koretz, Stecher, and Deibert in their 1992 study of Vermont's assessment system.

Because each alternative assessment program will have different parameters and will be managed from a different level within the system, it is important to consider all of the levels identified in Figure 1, as well as the different expenditure components identified there. Identifying the time spent by different personnel that is not directly compensated requires a complete understanding of all individuals who might bear some form of the cost. This could include administrators at the state, district, and school level, teachers who are required to participate in new assessment programs and be trained in their operation, and students for whom more time may be taken away from

instruction and devoted to assessment. The student costs must also be considered in light of any additional instructional benefits that accrue as a result of the new assessment program.

Costs Allocated to Students

The purpose of an assessment system is to gather more information on student performance and get a better sense of how much students are learning. One of the arguments made in favor of a number of alternative assessment tools is that they can work with the instructional program to improve student learning and performance. While it is not the purpose of this paper to discuss how successful these programs are, it is important to keep in mind the costs borne by students as assessment methods are changed.

The most obvious cost to students occurs if substantial training of individual teachers is required and takes place during school time. In that situation, the students can either be released from school (a difficult proposition in most states), or substitute teachers can be provided. To the extent that student performance is retarded by the absence of the regular classroom teacher, students bear a portion of the costs of a new assessment program. On the other hand, if the new program works with the curriculum in ways that improve overall student achievement, the effect is negated.

A second cost that could accrue to students would be due to instructional time lost because of additional time being devoted to assessments. Unless the new assessment program completely replaces the existing system, which seems unlikely upon initial implementation, additional time will be spent in assessment activities. If this leads to improved student outcomes, then there is a net gain to the student; if the opposite happens, then there is a cost.

Measuring these costs to students is probably an impossibility. The most obvious difficulty is determining how much learning is forgone due to the lost teacher time and increased assessment time, if any. Moreover, assigning benefits due to gains in student performance is equally difficult. This is a significant issue since many advocates of new assessment practices argue that good assessment serves an instructional function and should not even be considered as time away from instruction.

Finally, measuring gains or losses in student achievement may be impossible as a switch to a new assessment program is made. It seems safe to assume that a new assessment program would include different measures of student performance, and student progress in the new system might not be directly comparable to progress in the old system. It is even conceivable that under a new assessment system tied to a new instructional methodology, the old and new assessment systems would yield vastly different measures of success. Thus, estimation of the costs and/or benefits to students would be very difficult.

Overall, estimating the costs or benefits of an alternative assessment system for students is a very complex and difficult process. Assuming it is possible to reach agreement as to how the costs and benefits accruing to students could be measured, comparing those costs or benefits across alternatives, or with the benefits of other educational programs could be very difficult. Up to this point, we have attempted to use the monetary value of the resources allocated to the activity to reach a standardized measure of the expenditures or costs. Such a monetary value is much more difficult to estimate for students. Since our measure of costs or benefits for students is based on student achievement and not on time spent in alternative activities, it might be possible to use some measure of achievement gain per unit of time to estimate benefits, but it is still difficult to place a value on that. Furthermore, one needs to decide which assessment option should be used to measure that progress.

The Effect of Time

To this point, the issue of time has not been considered. Yet the effect of time on an evaluation of an alternative assessment program is critical. Again, assuming one of the goals of alternative assessment programs is to improve student performance through a link with the instructional program, then consideration of time is crucial to a cost analysis. As the discussion in section 3 shows, many of the expenditures incurred on behalf of an alternative assessment program will occur in the early years of the program as new assessment instruments are developed and school staff are trained in their use. Over time, these costs will diminish, and the annual operating costs will decline. To the extent that the assessment practices are woven into teacher training programs, and the development of new assessment items becomes a standard process, these costs will be lower.

If the assessment program is successful, one would anticipate many of the benefits derived from its implementation would increase over time. As teachers become more skillful at linking the assessment to instructional programs, long-term student outcomes would, it is hoped, improve. This would lead to greater benefits over time. The effect of this is that there is an initial high start-up cost, followed by reduced annual costs. At the same time, the initial benefits of the program will be lower than the long-term benefits. Unfortunately, this hypothesis can not be tested with cross sectional data, or even with only one or two years of implementation and assessment data. Rather, it will be necessary to conduct multiyear longitudinal studies of assessment programs to ascertain whether or not long-term costs decline and long-run benefits increase. If this hypothesis is true, then the longer the time frame considered in the cost analysis, the greater the probability that the benefits of a new assessment program will exceed its costs. Thus, in the short term, investment in new assessment practices may not seem as effective as such investment will appear in the longer term.

Summary

This section has looked more closely at the kinds of costs that are harder to evaluate than the expenditures discussed in the third section. The costs of personnel time devoted to an alternative assessment program, but not compensated directly through that program, must be evaluated in some consistent way. The most direct is to value the time spent on the basis of the wages paid the individuals engaged in the assessment activity. The difficulty occurs in determining who actually bears the cost of an individual's activities on behalf of the new program. The individual could bear a portion of the costs, as could the school community in general, or the students in the classes that are affected by the new assessment system.

Estimating the costs and/or benefits that accrue to students is more difficult as a straightforward monetary metric is not as easily developed. Moreover, assessment of student performance may vary depending on the assessment instrument chosen. Finally, the element of time is critical to a good cost analysis. Alternative assessment programs may have high initial costs, with gains in student performance tied to the assessment program not appearing

until some time in the future. Consequently, the time horizon selected becomes critical to the outcome of the analysis of the cost effectiveness of the program.

5. SUMMARY AND CONCLUSIONS

Estimating the costs of alternative assessment programs is a complex and difficult problem. The conceptual framework developed in this paper shows that there are a number of important issues to be resolved before accurate cost estimates can be developed. The first of these issues is developing a clear understanding of the difference between expenditures and costs. Expenditures represent the monetary value of the resources necessary to implement an assessment program, while the costs represent the value of the benefits of the best alternative that is forgone in establishing the assessment program.

The first step in this framework is to identify the expenditures necessary to operate an assessment program. This can be done by listing the resources necessary to make the program work, and then calculating the costs of those resources. The estimation of the expenditures necessary for all of the resources is relatively straightforward; however, identifying all of the components of expenditures for an assessment program, along with the levels at which those expenditures occur and the kinds of resources utilized, requires a great deal of work.

Once this task is complete, however, the expenditure estimate may not be an accurate reflection of the total costs of the alternative assessment program. A number of complex cost considerations must be resolved before the cost estimate is complete. These considerations include determining the value of personnel time devoted to the assessment program but not directly compensated by the program. Administrators, teachers and students who participate in various ways in the alternative assessment program will bear a share of the total costs. Estimating the value of this time is difficult given that knowledge about the alternative uses of that time by the individuals involved may not be available or may not be accurate.

In addition, a number of other cost components must be considered including the costs of operating both the new alternative assessment program and the previous assessment program, compared to the costs of having one

replace the other. In either case, there are considerable forgone benefits to be considered. In the case of total replacement, comparing the results of the new assessment with the old ma be impossible. Moreover, gaining the cooperation of affected parties may require extensive efforts and entail considerable costs. Again, the difficult issue here is to identify all of the forgone benefits and determine accurately all of the resources being devoted to the assessment program.

The benefits of many educational programs are often difficult to measure, particularly when there are multiple, and possibly conflicting, program goals. Under these circumstances, cost analysts in education have often resorted to estimating the monetary value of the resources devoted to the program being evaluated. In the end, this is probably the best way to estimate the costs of alternative assessment programs in educational settings. However, it is important to remember the opportunity costs that result from time commitments of individuals not directly compensated through the assessment program, such as teachers who are required to spend time on tasks that previously did not exist or were not their responsibility. Determining the value of these opportunity costs will improve the quality of educational cost analyses dramatically.

REFERENCES

Herman, J. L. (1992). What research tells us about good assessment. *Educational Leadership, 49*(8), 74-78.

Koretz, D., Stecher, B., & Deibert, E. (1992). *The Vermont Portfolio Assessment Program: Interim report on implementation and impact, 1991-92 school year* (CSE Tech. Rep. No. 350). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Levin, H. M. (1983). *Cost effectiveness: A primer.* Beverly Hills, CA: Sage Publications.

Monk, D. H. (1993). *The cost of systemic education reform: Conceptual issues and preliminary estimates.* Typescript.

O'Neil, J. (1992). Putting performance assessment to the test. *Educational Leadership, 49*(8), 14-19.

Picus, L. O. (1993a). *The allocation and use of educational resources at the* district *level: Evidence from the Schools and Staffing Survey* (Working Paper No. 33). Los Angeles: University of Southern California, Center for Research in Educational Finance.

Picus, L. O. (1993b). *The allocation and use of educational resources at the* school *level*: *Evidence from the Schools and Staffing Survey* (Working Paper No. 37). Los Angeles: University of Southern California, Center for Research in Educational Finance.

Shavelson, R. J., & Baxter, G. P. (1992). What we've learned about assessing hands-on science. *Educational Leadership, 49*(8), 20-25.

Shavelson, R. J., Gao, X., & Baxter, G. P. (1993). *Sampling variability of performance assessments* (CSE Tech. Rep. No. 361). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Wiggins, G. (1992). Creating tests worth taking. *Educational Leadership, 49*(8), 26-33.